

University of Mississippi

eGrove

Faculty and Student Publications

Psychology

7-15-2019

Too much model, too little data: How a maximum-likelihood fit of a psychometric function may fail, and how to detect and avoid this

Nicolaas Prins
University of Mississippi

Follow this and additional works at: https://egrove.olemiss.edu/psych_facpubs

Recommended Citation

Prins, N. (2019). Too much model, too little data: How a maximum-likelihood fit of a psychometric function may fail, and how to detect and avoid this. *Attention, Perception, & Psychophysics*, 81(5), 1725–1739. <https://doi.org/10.3758/s13414-019-01706-7>

This Article is brought to you for free and open access by the Psychology at eGrove. It has been accepted for inclusion in Faculty and Student Publications by an authorized administrator of eGrove. For more information, please contact egrove@olemiss.edu.



Too much model, too little data: How a maximum-likelihood fit of a psychometric function may fail, and how to detect and avoid this

Nicolaas Prins¹

Published online: 15 March 2019
© The Psychonomic Society, Inc. 2019

Abstract

Maximum-likelihood estimation of the parameters of a psychometric function typically occurs through an iterative search for the maximum value in the likelihood function defined across the parameter space. This procedure is subject to failure. First, iterative search procedures may converge on a local, not global, maximum in the likelihood function. The procedure also fails when the likelihood function does not contain a maximum. This is the case when either a step function or a constant function is associated with a higher likelihood than the model can attain with finite parameter values. In such cases iterative search procedures may erroneously report having successfully converged on a maximum in the likelihood function. This will lead not only to inaccurate models for the observed data, but may also lead to inaccurate results regarding the reliability of parameter estimates, goodness-of-fit of the model, or model selection. I describe a method by which such false convergences can be reliably detected. I also present results of simulations that systematically investigate how stimulus placement, number of trials, parameters estimated, task (2AFC, 4AFC, etc.), and whether the lapse rate is allowed to vary affect the probability that the likelihood function will not contain a maximum. Based on the results of the simulations recommendations are made regarding experimental design and modeling choices. Software that implements the method is made available for downloading.

Keywords Model selection · Statistics · Statistical inference

Introduction

Behavioral responses in a psychophysical task are often modeled using some parametric function that allows the behavior to be characterized using just a few parameter values. These parameter values can then be compared between different experimental conditions in order to determine whether the experimental manipulation exerts an effect on performance. Performance in a psychophysical task is often measured as a binary response, for example with categories “yes” and “no” in a contrast-detection task or “correct” and “incorrect” in an m-Alternative Forced Choice Task (mAFC, e.g., Kingdom & Prins, 2016). In what follows, I use the terms “positive” and “negative” response for the two response categories in such tasks. A generic formulation of the psychometric function (PF) that relates the probability of observing a positive

response to a quantitative stimulus characteristic (e.g., stimulus intensity x) is given by:

$$\psi(x; \alpha, \beta, \gamma, \lambda) = \gamma + (1 - \gamma - \lambda)F(x; \alpha, \beta) \quad (1a)$$

where $F(x; \alpha, \beta)$ is typically a sigmoidal function with a range between 0 and 1 such as the logistic or Weibull function (e.g., Hall, 1981; Treutwein & Strasburger, 1999). In this paper, the Logistic function will be used exclusively (but the arguments and results in this paper generalize to any sigmoidal function). The Logistic function is given by:

$$F_{\text{logistic}}(x; \alpha, \beta) = \frac{1}{1 + \exp(-\beta(x - \alpha))} \quad (1b)$$

An example logistic PF that is fitted to some data is shown in the upper panel of Fig. 1.

The function F is assumed to describe the characteristics of the perceptual mechanism underlying the task (e.g., Kingdom & Prins, 2016). It is characterized by its specific shape (Logistic, Gumbel, etc.) and two parameters. Parameter α determines the location of the function while parameter β determines the steepness of the function. Parameter α is often referred to as the “threshold,” a term that betrays historical

✉ Nicolaas Prins
nprins@olemiss.edu

¹ Department of Psychology, University of Mississippi,
Oxford, MS 38677, USA

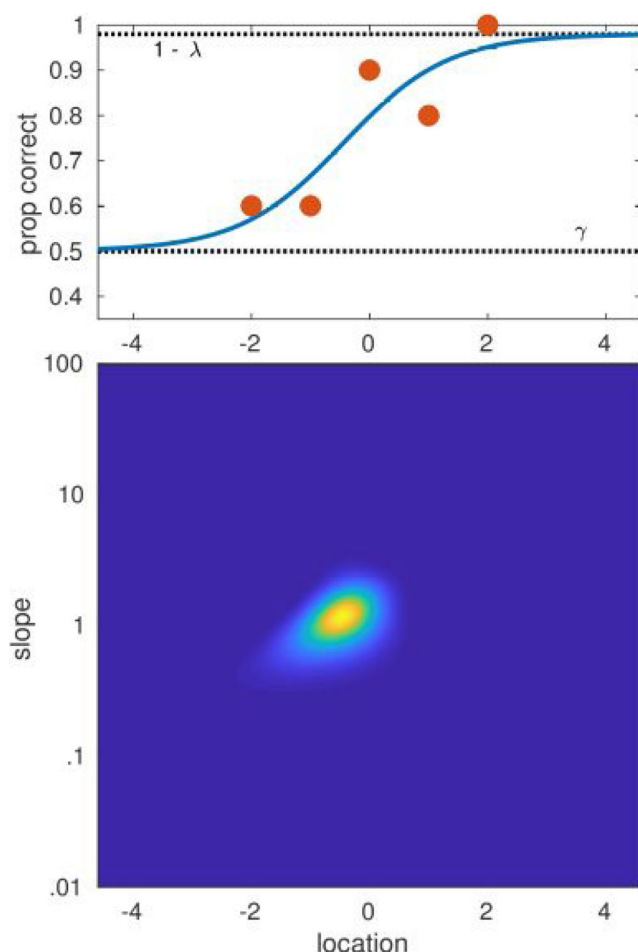


Fig. 1. **Top:** Some hypothetical data with a maximum-likelihood fit Logistic function. Each data point is based on ten trials. **Bottom:** Likelihood function across *location/slope* values

misconceptions regarding the nature of the underlying perceptual mechanism (e.g., Swets, 1961). Here, I refer to α as the *location* parameter, a term that describes a characteristic of the function itself and does not carry any implications regarding the mechanism assumed to underlie it. I refer to parameter β as the *slope* parameter. The parameters γ and λ in equation (1) do not characterize the perceptual process. Instead, parameter γ determines the lower asymptote of the function and would correspond to the false alarm rate in a “yes/no” task or to the reciprocal of the number of response alternatives in an mAFC task. Because of the latter, parameter γ is often generically referred to as the *guess rate*. While under the assumptions of signal detection theory an observer never truly guesses, if an observer *were* to guess in an mAFC task, the probability of a correct response would be equal to the reciprocal of the number of response alternatives in an mAFC task. Parameter λ determines the upper asymptote of the function and is determined by the rate at which the observer makes stimulus-independent responses (e.g., due to attentional lapses or finger errors). Parameter λ is for that reason often referred to as the *lapse rate*. Because researchers are, with few exceptions (e.g.,

van Driel et al., 2014), only interested in F and its parameters α and β , the remaining two parameters are often referred to as nuisance parameters.

Commonly, the best-fitting PF to some data is found using maximum-likelihood estimation of parameter values. The likelihood associated with a PF with parameter values $\alpha = a$, $\beta = b$, $\gamma = g$, and $\lambda = l$, is defined as:

$$L(a, b, g, l | \mathbf{y}) = \prod_{k=1}^N p(y_k | x_k; a, b, g, l), \quad (2)$$

where $p(y_k | x_k; a, b, g, l)$ gives the probability of observing response y (1 for a positive response, 0 for a negative response) on trial k given stimulus intensity x_k and PF parameter values a , b , g , and l . Of course,

$$p(y_k | x_k; a, b, g, l) = \psi(x_k; a, b, g, l) \text{ if } y_k = 1 \text{ and}$$

$$p(y_k | x_k; a, b, g, l) = 1 - \psi(x_k; a, b, g, l) \text{ if } y_k = 0.$$

While strictly almost certainly false, the above assumes that the parameter values remain stable throughout the experiment. An example violation of this assumption would be participant fatigue which may lead to an upward drift in location parameter or a decrease in slope parameter across the experiment. The above also assumes that responses are independent. An example violation of this assumption would be an observer being less likely to respond “no” because a high number of consecutive “no” responses preceded the current trial. Alternative models may be fitted to accommodate violations of independence and stability (e.g., Fründ, Haenel, & Wichmann, 2011).

During the fitting procedure, any of the four parameters may be fixed at a constant value while the remaining are free to vary. For example, the bottom panel of Fig. 1 gives the likelihood function for the model in equation (1) across values for the *location* and *slope* for the data shown in the top panel. The *guess* and *lapse rate* were fixed at values of 0.5 and 0.02, respectively. In this example, the likelihood function displays a single maximum. Using the maximum-likelihood criterion, the values for the *location* and *slope* at this maximum define the best-fitting PF.

Generally speaking, the maximum in the likelihood function cannot be determined using analytical methods and must instead be found using numerical approximation. A common algorithm that is used to locate the maximum in the likelihood function is Nelder and Mead’s (1965) simplex method. The simplex method can be used to locate a maximum in parameter spaces of any dimensionality.¹ The simplex method performs well when the function to be

¹ Note that the simplex method as proposed by Nelder and Mead (1965) *minimizes* a function. To maximize a function, in order to find the maximum in the likelihood function for example, the method can be applied to the negative likelihood or, for practical purposes, to the negative log likelihood. In order to avoid awkward and repetitive phrasing, in this paper I write as if the Nelder-Mead method maximizes a function.

maximized is concave and contains a unique maximum (e.g., Kolda et al., 2003; Lagarias et al., 1998).

Local maxima in the likelihood function

In simple logistic or probit regression (i.e., when F in equation [1] is the logistic or the cumulative Gaussian and the upper and lower asymptote equal zero and one, respectively) it can be shown that any maximum in the likelihood function must be the global maximum (Pratt, 1981). However, barring these conditions, the likelihood function may contain local maxima. An example of this is shown in Fig. 2. Here the likelihood function contains a local maximum (at $\alpha = 0.677$, $\beta = 1.325$). The PF corresponding to this local maximum is shown in the top panel using the broken blue line. However, the likelihood function also contains a region of parameter values associated

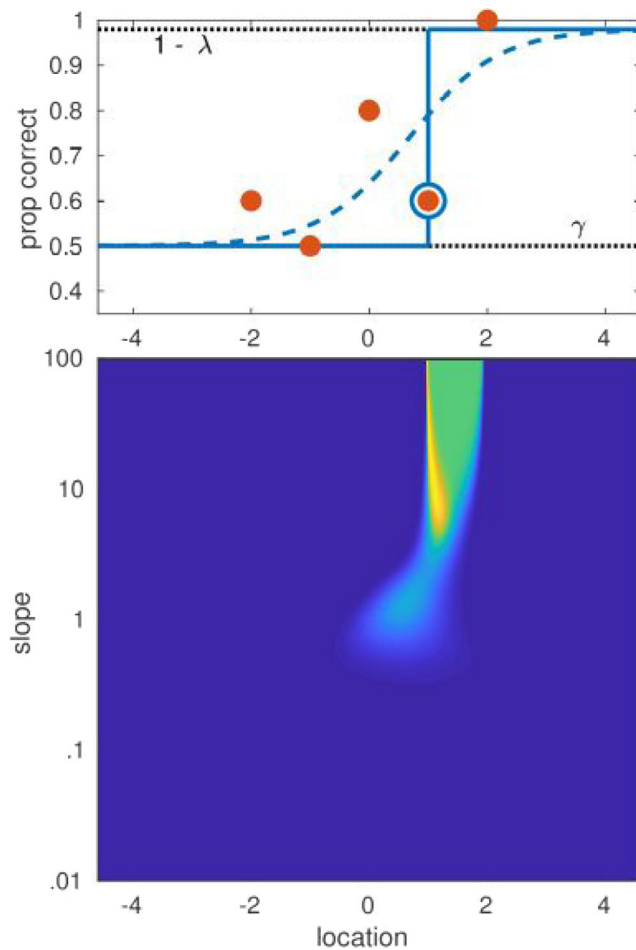


Fig. 2. The broken line in the top panel shows a function that is located at a local maximum in the likelihood function (shown in bottom panel). The step function shown has a higher likelihood than any sigmoidal PF with $\gamma = 0.5$ and $\lambda = 0.02$, but can be approached by such a function to any arbitrary degree of precision. Note that the exact color shown in the likelihood function image at high *slope* values and near *location* values equal to 1 (i.e., for functions that closely approximate the step function) is affected by aliasing in the image

with higher likelihoods. Nevertheless, the likelihood function does not contain a global maximum (defined as a point in parameter space that has a likelihood higher than any other point and for which all points in its immediate vicinity have lower likelihood). The step function shown in the top panel of the figure in solid blue can be approximated by a sigmoidal PF to any arbitrary degree of precision but has a higher likelihood than any sigmoidal PF. Below I discuss such a scenario in more detail. For now, it should be noted that unless preventative measures are taken, it is quite possible for the Nelder-Mead Simplex search to converge on the local maximum and falsely identify it as the maximum-likelihood solution. It may also be noted that by visual inspection of the fit to the data, it is not at all obvious that the local maximum would correspond to anything other than a global maximum. It would also be not at all obvious from a visual inspection that the step function in the figure has a higher likelihood than the local maximum in broken blue. Moreover, it is far from obvious that the step function shown is the highest likelihood function that can be approached by a sigmoidal PF. All this despite the fact that this would be about the easiest design in which to detect these problems (i.e., a mere five equally-spaced and equally-sampled stimulus intensities were used). In a later section, I discuss how it can be determined that the step function shown has a higher likelihood than any other function that may be approached by a sigmoidal function given the constraints on the *guess* and *lapse rates*.

In order to avoid identifying a local maximum as the global maximum, it is important to ensure that an appropriate starting position for the iterative search is chosen. One strategy to accomplish this is to perform a brute-force search through a discrete parameter grid, then use the function in the grid that is associated with the highest likelihood as the seed for the iterative search procedure.² As long as the range of parameter values included in the brute-force grid encompasses the global maximum and the grain of the grid is not excessively coarse, this method will successfully converge on the global maximum in the likelihood function if indeed it exists. While it may appear inefficient to search through a large number of PFs contained in a grid, in practice it will often actually result in reduced fitting times compared to using an arbitrary seed. The calculations performed during the brute-force grid search can be vectorized while the serial iterative search procedure

² Another commonly used strategy is to run multiple iterative searches, each started at random or systematically varied positions in parameter space. The solution that is converged on by several such searches is taken to correspond to the true maximum in the likelihood function. The searches that resulted in convergence failures or inconsistent solutions would be considered to be due to poor starting positions. Note, however, that this strategy would actually identify the local maximum as the solution in scenarios such as those displayed in Fig. 2. This is because only the searches that move toward the local maximum will converge. While some of the searches that move toward the step function may falsely report convergence, their parameter estimates would not be consistent.

can not. Starting the iterative search near the maximum likelihood solution, as opposed to an arbitrary position in parameter space, significantly reduces the number of iterations needed to reach convergence.

No global maximum exists in the likelihood function

Figures 3, 4, and 5 contain three different hypothetical results and likelihood functions of an experiment. Each of these experiments consists of 50 trials, ten trials at each of five different (log) stimulus intensities, equally spaced between -2 and 2. Each of these results was produced by a simulation in which the generating function was the Logistic function with $\alpha = 0$,

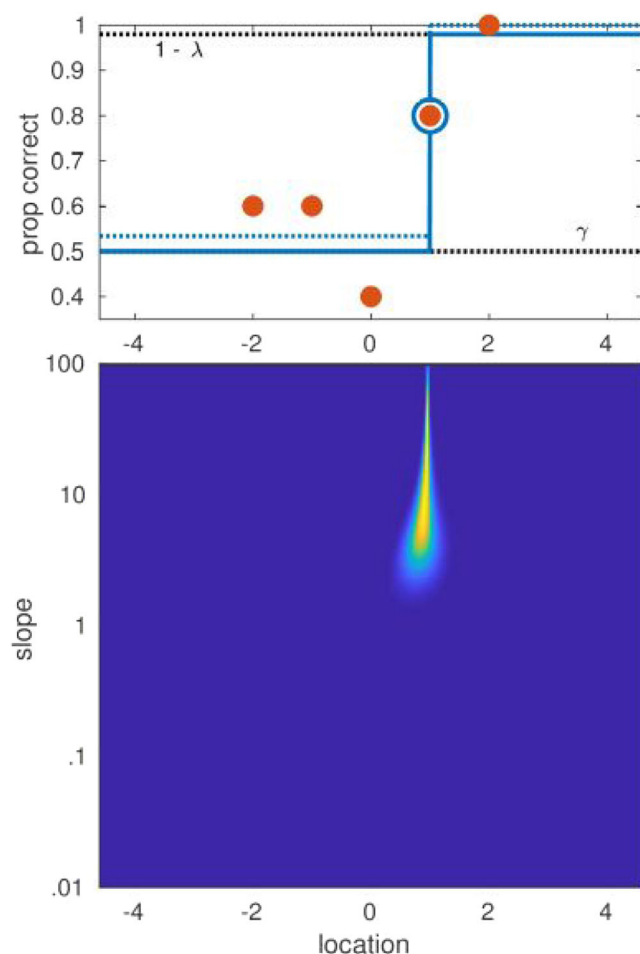


Fig. 3. Example of Scenario -1. The step function shown has a higher likelihood than any sigmoidal PF (with $\gamma = 0.5$ and $\lambda = 0.02$) but can be approached by such a function to any arbitrary degree of precision. If the *guess rate* is freed the step function will have the broken line left of the discontinuity as the lower asymptote. If the *lapse rate* is a free parameter the step function will have the broken line right of the discontinuity as the upper asymptote. Note that the exact color shown in the likelihood function image at high *slope* values and near *location* values equal to 1 (i.e., for functions that closely approximate the step function) is affected by aliasing in the image

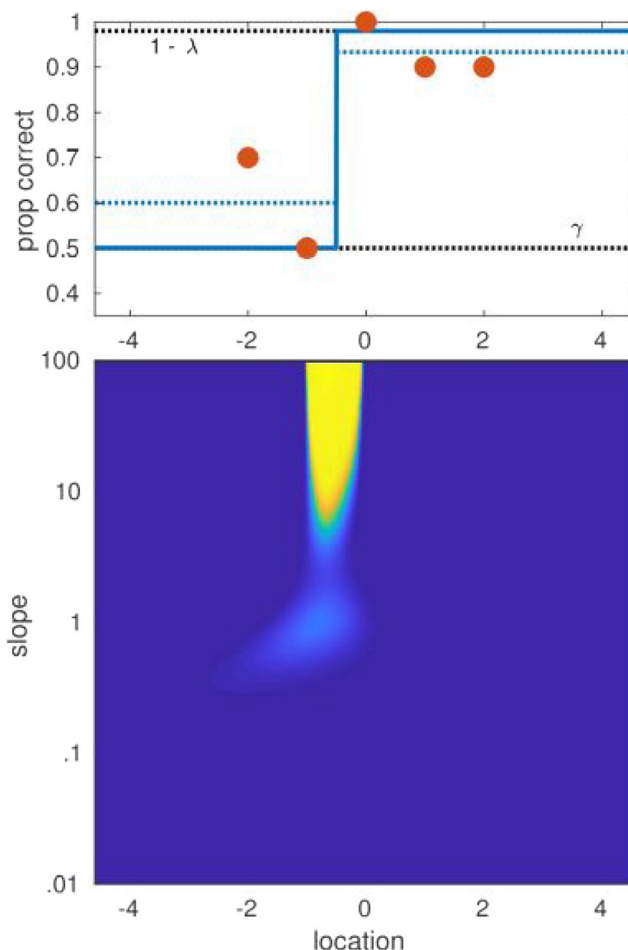


Fig. 4. Example of Scenario -2. The step function shown has a higher likelihood than any sigmoidal PF (with $\gamma = 0.5$ and $\lambda = 0.02$) but can be approached by such a function to any arbitrary degree of precision. If the *guess rate* is freed the step function will have the broken line left of the discontinuity as the lower asymptote. If the *lapse rate* is freed the step function will have the broken line right of the discontinuity as the upper asymptote

$\beta = 1$, $\gamma = 0.5$, and $\lambda = 0.02$. This function evaluates to 0.577, 0.629, 0.740, 0.851, and 0.923 at the five stimulus intensities, respectively. For now, I discuss fitting these three datasets using fixed values for the lower asymptote (γ) and higher asymptote ($1 - \lambda$). Later I generalize to fits that free the lower and/or higher asymptote.

The likelihood functions for the simulated experiments shown in Figs. 3, 4, and 5 do not contain global maxima. The absence of a global maximum in the likelihood function may happen in circumstances that I group into three scenarios exemplified by the figures. In all three scenarios, a function that can be approached to any arbitrary degree of precision by the standard functions that are used to model PFs (e.g., Logistic, Weibull, or any other increasing and continuous sigmoidal function) provides a greater likelihood than can be attained by the sigmoidal function itself. In practice, during a fitting procedure an iterative search procedure such as

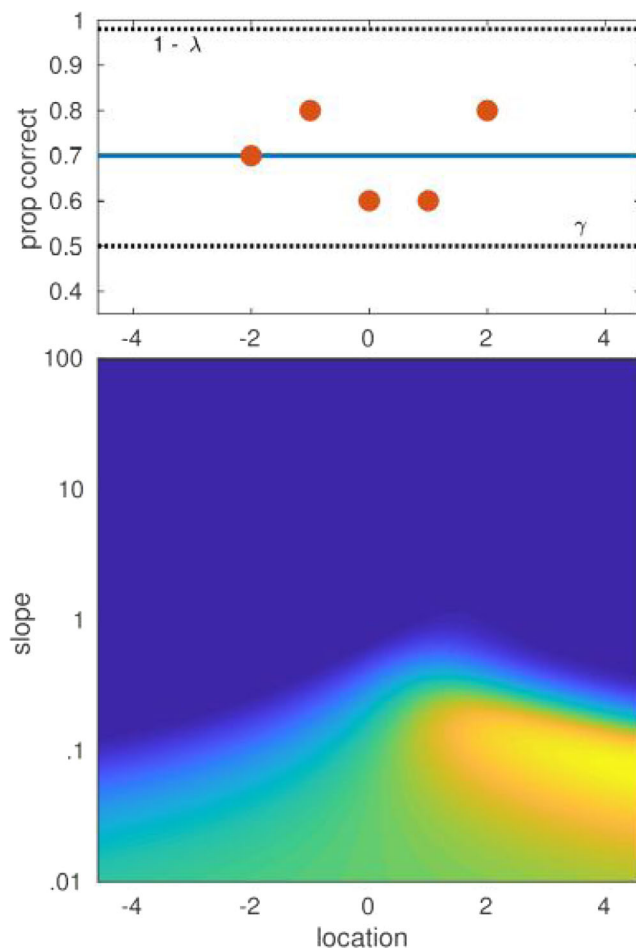


Fig. 5. Example of Scenario -3. The constant function shown has a higher likelihood than any increasing sigmoidal PF (that has the value 0.7 in its range while the *guess* and/or *lapse rate* may be free parameters or not) but can be approximated by such a function to any arbitrary degree of precision

Nelder and Mead's (1965) simplex search will chase after one of these functions but will never exactly attain it. In the scenarios shown in Figs. 3 and 4 the search procedure will approximate (but never exactly match) a step function. For the scenario shown in Fig. 5 the search procedure will approximate (but never match) a constant function. In all of these three cases, the procedure should not locate a maximum in the likelihood function (since none exists) and should instead signal that convergence failed after the allowed number of iterations or function evaluations has been reached. In the Appendix, I present an argument that the lack of a maximum in the likelihood implies that a step function or constant function that can be approached by a PF must exist that has higher a likelihood than such a PF.

However, in practice, the search procedure may signal convergence on a (non-existing) maximum. This will happen whenever the criterion (or criteria) used to signify convergence has been satisfied even though a maximum has not actually been reached. The probability with which an iterative

search procedure will falsely signal convergence depends in a non-trivial manner on the specifics of the convergence criteria as well as the specific data set. A reported fit that resulted from a false convergence might appear as a reasonable fit and it may not be at all obvious that a step or constant function might provide a better fit than a PF with some finite-valued positive *slope* (recall the example in Fig. 2). It will be especially difficult to detect false fits in designs in which trials are unevenly distributed across stimulus intensities that are not equally-spaced or when each trial utilizes a unique stimulus intensity as might occur in adaptive designs. Moreover, while researchers might reasonably be expected to inspect the reported fit to some human- or animal-generated data, it is not a reasonable expectation that each simulation is visually inspected when the reliability of parameter estimates is established through a bootstrapping procedure. When false convergences occur in the fits to bootstrapped samples for which in fact no global maximum exists, the result will be that the imprecision in the parameter (as judged, for example, by the standard deviation of the bootstrapped estimates; Efron & Tibshirani, 1993) is underestimated, resulting in an inflation of type-I error rates. Similarly, when model comparisons (including determination of Goodness-of-Fit) are performed using Monte Carlo simulations the diagnostics will be systematically misestimated if false convergences occur.

Scenario -1

In Fig. 3, the step function shown as the solid blue line has a higher likelihood than any increasing sigmoidal function that is constrained to have lower asymptote $\gamma = 0.5$ and upper asymptote $1 - \lambda = 0.98$. Note that, for any finite value of *slope* β a value of *location* α exists for which $\psi(x = 1) = 0.8$ (the observed proportion correct at $x = 1$). The higher the value of β is in such a combination of α and β values, the better the resulting function will fit the remaining stimulus intensities overall. Thus, during an iterative search procedure to fit the PF, the estimate for *location* parameter α approaches the utilized stimulus intensity $x = 1$ while the estimate for *slope* β approaches $+\infty$. As a result, the function will approach a step function such as that shown in the figure. I will express this as:

$$\lim_{\alpha \rightarrow x_s, \beta \rightarrow \infty, \gamma = g, \lambda = l} \psi(x) = \begin{cases} g, & \text{if } x < x_s \\ \frac{nC(x_s)}{n(x_s)}, & \text{if } x = x_s \\ 1-l, & \text{if } x > x_s \end{cases} \quad (3)$$

where subscript s enumerates the different stimulus intensities (low to high), g is the fixed value of the *guess rate*, l is the fixed value for the *lapse rate*, $nC(x_s)$ is the number of correct responses observed at intensity x_s , and $n(x_s)$ is the number of trials presented at intensity x_s . The likelihood associated with this step function is calculated as:

$$L = \left(\frac{nC_s}{n_s}\right)^{nC_s} \times \left(1 - \frac{nC_s}{n_s}\right)^{(n_s - nC_s)} \\ \times \prod_{i < s} \left(\gamma^{nC_i} \times (1 - \gamma)^{(n_i - nC_i)}\right) \\ \times \prod_{i > s} \left((1 - \lambda)^{nC_i} \times \lambda^{(n_i - nC_i)}\right) \quad (4)$$

This scenario will be assigned the arbitrary numerical code -1 (a negative number is used to indicate that no maximum in the likelihood function exists and that technically the fit failed).

Scenario -2

In Fig. 4, the step function shown as the solid blue line has a higher likelihood than any increasing sigmoidal function that has lower asymptote $\gamma = 0.5$ and upper asymptote $1 - \lambda = 0.98$. During an iterative search procedure to fit the PF, the estimate for *location* parameter α will assume some value in the interval (x_s, x_{s+1}) while the estimate for *slope* β will approach $+\infty$ ($s = 2$ in the example in Fig. 4). The resulting function will approach a step function such as that shown in the figure. I will express this as:

$$\lim_{x_s < \alpha < x_{s+1}, \beta \rightarrow \infty, \gamma = g, \lambda = l} \psi(x) = \begin{cases} g, & \text{if } x \leq x_s \\ 1 - l, & \text{if } x \geq x_{s+1} \end{cases} \quad (5)$$

The likelihood associated with this step function is calculated as:

$$L = \prod_{i \leq s} \left(\gamma^{nC_i} \times (1 - \gamma)^{(n_i - nC_i)}\right) \\ \times \prod_{i > s} \left((1 - \lambda)^{nC_i} \times \lambda^{(n_i - nC_i)}\right) \quad (6)$$

This outcome is referred to here arbitrarily as Scenario -2.

Scenarios -1 and -2 are akin to what are referred to respectively as “quasi-complete separation” and “complete separation” in the context of logistic regression (Albert & Anderson, 1984). However, in strict logistic regression the lower and upper asymptote are assumed to equal 0 and 1, respectively. Under these assumptions quasi-complete or complete separation will only occur when all observed proportions below x_s equal 0 and all observed proportions above x_s equal 1. When the lower asymptote and/or upper asymptote is assumed to equal some value other than 0 or 1, respectively, (as is the case in the examples given here) quasi-complete separation is much harder to detect as it may occur in circumstances where observed proportions do not equal 0 or 1.

Scenario -3

In Fig. 5, the constant function shown as the solid blue line has a higher likelihood than any increasing sigmoidal function that has lower asymptote $\gamma = 0.5$ and upper asymptote $1 - \lambda$

$= 0.98$. During an iterative search procedure to fit the PF, the estimate for *location* parameter α will approach either $-\infty$ or $+\infty$ (depending on whether the overall observed proportion correct is, respectively, greater or less than the value of the PF at its *location* parameter [i.e., $\psi(x = \alpha)$]) while the estimate for *slope* β will approach 0. The specific combination of values for α and β will be such that the PF will approach a horizontal line at the overall proportion correct across all trials in the experiment. I will express this as:

$$\lim_{\alpha \rightarrow -\infty \text{ or } +\infty, \beta \rightarrow 0, \gamma = g, \lambda = l} \psi(x) = \begin{cases} g, & \text{if } pC \leq g \\ pC, & \text{if } g < pC < 1 - l \\ l, & \text{if } pC \geq 1 - l \end{cases} \quad (7)$$

where $pC = \frac{\sum_i nC(x_i)}{\sum_i n(x_i)}$. That is, the constant function has a value equal to the overall proportion correct across all trials (provided this proportion correct has a value between g and $1 - l$). The likelihood associated with the constant function is given as:

$$L = pC^{\sum_i nC(x_i)} \times (1 - pC)^{\sum_i n(x_i) - \sum_i nC(x_i)} \quad (8)$$

This outcome is referred to here arbitrarily as Scenario -3. Note that I have taken some liberty in the expression above in that α and β need to covary (in a manner that is specific to the particular sigmoidal curve [logistic, cumulative normal, etc.] being fitted) in order to approximate a constant function that has the value pC . Nevertheless, for any sigmoidal function with parameters $\alpha = a$ and $\beta = b$ whose range includes pC , parameter values $\alpha = a^*$ and $\beta = b^*$ can be found for which the function more closely approximates the constant function at pC . The value for a^* will be closer to infinity (or negative infinity) than a is and the value for b^* will be closer to 0 than b is.

All scenarios may also occur when the *guess rate* and/or *lapse rate* are free parameters. The broken blue lines in Figs. 3 and 4 show the best fitting functions when both the *guess rate* and the *lapse rate* are free parameters. The solid blue line in Fig. 5 is the best fitting function whether the *guess* and/or *lapse rates* are free to vary or not. Again, the functions shown in Figs. 3, 4, and 5 are functions that can be approximated to any degree of precision by a sigmoidal PF, but can never be attained exactly. For the step functions in Figs. 3 and 4 (Scenario = -1 and -2 respectively), freeing the *guess rate* results in an estimate for the *guess rate* that is equal to the overall proportion correct across the trials at intensities below the discontinuity. For Scenario -1:

$$\psi(x_i) = \gamma = \frac{\sum_{i < s} nC(x_i)}{\sum_{i < s} n(x_i)}, \quad \text{for } x < x_i, \quad (9)$$

Note that the limits of the summation should read $i \leq s$ when Scenario = -2. Similarly, the estimate for the *lapse rate*

will be such that the height of the upper asymptote is equal to the overall proportion correct across the trials at intensities above the discontinuity:

$$\psi(x_i) = 1 - \lambda = \frac{\sum_{i>s} nC(x_i)}{\sum_{i>s} n(x_i)}, \quad \text{for } x > x_i \quad (10)$$

Note that freeing the *lapse rate* and/or *guess rate* will not necessarily result in a fit that has the same *location* of the discontinuity in Scenarios -1 and -2. It may also be the case that freeing the *guess* and/or *lapse rate* will result in a fit that falls in a different scenario altogether.

Provided that the overall observed proportion correct lies within the interval $(\gamma, 1 - \lambda)$, fixing the *slope* parameter (β) at a constant value will prevent Scenarios -1, -2, and -3. Of course, the value of the *slope* parameter is often of theoretical significance in which case fixing its value would defeat the purpose of the experiment. If we loosen the definition of the maximum of a function and allow the limiting functions in equations (3), (5), and (7) to be regarded as the maximum likelihood fit, we can derive parameter estimates (even though not all will be finite-valued). For example, in Scenarios -1 and -2, where the limiting function is a step function, the true, generating *location* parameter value is likely near the discontinuity in the step function and the location of the discontinuity can serve as an estimate of the *location* parameter even though the likelihood function does not contain a true global maximum.

Identifying the scenario of a fit

As mentioned above, even if the iterative search reports that a maximum has been found, it may have done so in error. In case it reports that it has failed to converge on a maximum, it may not be clear to which scenario the results adhere (recall again the example in Fig. 2). However, it is relatively straightforward to determine whether one of the limiting functions given above fits the data better than the fit resulting from the iterative search. A key consideration is that when the scenario is -1, -2, or -3, the best-fitting limiting function can be identified without uncertainty. It does not require an iterative search for parameter values. For any data set that used k different stimulus intensities there are at most k possible fits that fall in Scenario -1, $k - 1$ possible fits that fall in Scenario -2, and only one possible fit that falls in Scenario -3. All that needs to be done is to perform a brute force search through the likelihoods of the $2 \times k$ possible fits in the categories -1, -2, and -3 using the limiting functions given above. These likelihoods can then be compared to the likelihood that resulted from the iterative search. The function associated with the highest likelihood is the best-fitting function that can be attained (if Scenario = 1) or approximated (else).

Assignment of values to parameter estimates

For results that fall in Scenario 1, assignment of parameter values is straightforward. There is a maximum in the likelihood function and the corresponding parameter values are finite. As mentioned earlier, even when the likelihood function does not contain a strict maximum, the data still contain information as to the value of the *location* or *slope* parameters. For fits resulting in Scenario -1, the *location* estimate may be defined to correspond to the value it approaches: the discontinuity (x_s in equation 3). Similarly, the *slope* estimate may be defined as the value it approaches: $+\infty$. If the *guess* and/or *lapse rate* are free parameters, their estimated values will be as described above (equations 9 and 10).

For fits resulting in Scenario -2, it is less clear what exact value might be assigned to the *location* parameter. It is clear that for the limiting step function any location for the discontinuity in the interval (x_s, x_{s+1}) leads to equal likelihoods and we may, somewhat arbitrarily, define the *location* parameter estimate to have a value of $\frac{x_s + x_{s+1}}{2}$ in Scenario -2. The *slope* estimate will be assigned the value of $+\infty$. If the *guess* and/or *lapse rate* are free parameters, their estimated values will be as described above (equations 9 and 10).

When a fit results in Scenario -3, the *slope* estimate may be assigned a value of 0 (the value it approaches). If the observed overall proportion correct is greater than the value of the PF at its *location* parameter, $\psi(x = \alpha)$, the *location* parameter is assigned the value of $-\infty$ (the value it approaches during the iterative search). Otherwise it is assigned the value of $+\infty$ (ditto). There are a few exceptions to this rule. If the overall proportion correct is greater than or equal to the upper asymptote, a constant function within the stimulus range can also be approached by a PF with a near infinite *slope* and any *location* value less than the lowest stimulus intensity used. Similarly, if the overall proportion correct is less than or equal to the lower asymptote the constant function within the stimulus range can be approached by a PF with a near infinite *slope* and any *location* value above the highest stimulus intensity used.³ In case the *guess* and/or *lapse rate* are free parameters and the fit results in Scenario -3, the constant function within the stimulus range can also be accomplished by a step function. Thus, in these cases none of the free parameters can reasonably be assigned values in this scenario.

Note that a value of 0 or $+\infty$ for the *slope* parameter or a value of $+\infty$ or $-\infty$ for the *location* parameter are all biologically extremely implausible. Thus, the true, generating parameter cannot correspond exactly to those assigned in scenarios -1, -2, and -3. For that reason, assigning these values to parameters may strike some as inherently incorrect. However, maximum-likelihood fitting is entirely data-driven and is inherently not concerned with the plausibility of the resulting estimates. A Bayesian approach

³ I thank an anonymous reviewer for this insight.

does allow plausibility considerations to be incorporated into the estimates and will receive a bit more consideration in the *Discussion*. For now, one must remember that the maximum-likelihood procedure does not claim to find the parameter values that are most likely (or indeed even likely at all) to be correct. Instead, the maximum-likelihood procedure claims to find the parameter values that are most likely to reproduce the observed data set. Thus, within the confines of the maximum-likelihood framework, the proposed assigned values can be regarded as the maximum-likelihood estimates, regardless of their plausibility. In this sense, fits resulting in scenarios -1, -2, and -3 should be considered similarly to any other maximum-likelihood fit. That is, within the constraints of the model specifications, maximum-likelihood estimates correspond to those that give the highest probability of reproducing the observed data. No more, no less.

Simulations

The purpose of the following simulations is to elucidate how the design of an experiment affects the probability that the likelihood function contains a global maximum. Specific design choices that are investigated are the total number of trials (32, 64, 128, and 256), the stimulus placement strategy (see below), the number of response alternatives (two or four in an alternative-forced choice design), and whether the *lapse rate* is allowed to vary during fitting. I first discuss simulations in which the *lapse rate* was fixed followed by simulations in which the *lapse rate* was allowed to vary.

Simulations using a fixed lapse rate

The generating function in all simulations was the Logistic function with its *location* parameter (α) equal to 0 and its *slope* parameter (β) equal to 1. The *guess rate* (γ) varied as appropriate with the simulated design ($1/m$ in *m*AFC), while the *lapse rate* (λ) was always 0.02. All simulations were performed using the Palamedes toolbox (Prins & Kingdom, 2018). In each condition a total of 2,000 experiments were simulated.

Placement strategies

A variety of Method of Constant Stimuli (MOCS) placement strategies were used as well as an adaptive placement method that targets both the *location* and the *slope* parameter (the “psi method”; Kontsevich & Tyler, 1999). The different placement strategies are illustrated in Fig. 6. Two factors were varied in the placement design in MOCS strategies: The range of stimulus intensities covered and the number of stimulus intensities used. The range of stimulus intensities covered was expressed in terms of the range of values of the generating $F(x; \alpha, \beta)$ (see

equation 1). This range was 0.4 (“narrow”), 0.6 (“medium”), or 0.8 (“wide”). For all three ranges used, the stimulus intensities were equally spaced and symmetrical around the value of the generating *location* parameter. Either four or eight equally-spaced stimulus intensities were used. As an example, when four different stimulus intensities in the medium range (0.6) were used, the lowest stimulus intensity used was $F_{0.2}^{-1}$ (read: the intensity at which F evaluates to 0.2), the highest intensity used was $F_{0.8}^{-1}$, and the remaining two intensities were placed so as to create equal spacing (in terms of stimulus intensity) between intensities.

The psi adaptive method (Kontsevich & Tyler, 1999) updates a posterior distribution across a parameter space containing *location* parameter values and log-transformed *slope* values and places stimuli at an intensity that will maximize the expected information gained from the trial. Information is quantified as the (Shannon) entropy in the posterior distribution. In the simulations, the psi-method could choose from 31 stimulus intensities spaced equally between $F_{0.01}^{-1}$ and $F_{0.99}^{-1}$. The prior distributions were uniform within a constrained range of both *location* and log *slope* values. The base range for *location* values was $F_{0.01}^{-1}$ (-4.5951) through $F_{0.99}^{-1}$ (4.5951). This range of *location* values contained within the prior distribution was randomly jittered from the base range between simulated experiments. The jitter was taken from a uniform distribution with limits -0.6127 and 0.6127. The base range for log *slope* values was -2 through 2, randomly jittered by a value taken from a uniform distribution with limits -0.1333 and 0.1333. Each prior distribution contained 301 equally-spaced *location* values and 301 equally-spaced log *slope* values. The psi-method assumed the appropriate value for the *guess rate* (i.e., 0.5 for 2 AFC and 0.25 for 4 AFC) and a value of 0.02 (i.e., the generating value) for the *lapse rate*. In order to distinguish between the original psi method and some variations on the psi method that will be discussed later, I label the former as $\text{Psi}_{\alpha\beta}$ where the subscripted symbols indicate the parameters that were included in the posterior distribution.

All experiments simulated under $\text{Psi}_{\alpha\beta}$ were repeated as an MOCS experiment using the placement that resulted from the $\text{Psi}_{\alpha\beta}$ run. In these repeated experiments the responses were resampled from the generating function. This allows separating the effect from placement *per se* from the effect of the placement being adaptive. That is, without this additional condition it would be unclear as to whether any obtained advantage of $\text{Psi}_{\alpha\beta}$ placement is due merely to optimizing the distribution of trials across stimulus intensities with respect to the generating function or whether any advantage is due to the $\text{Psi}_{\alpha\beta}$ method being actively adaptive with respect to preceding trials in each run.

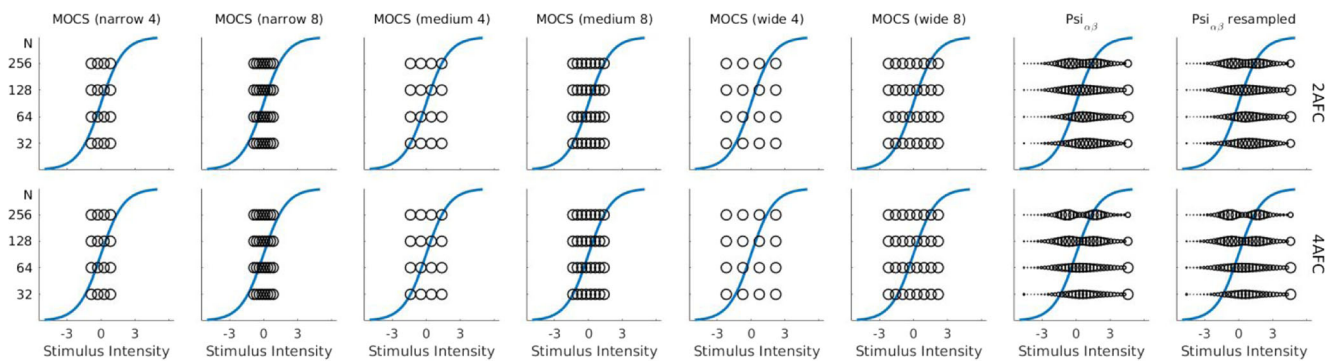


Fig. 6. Placements used in the simulations that were fit using a fixed *lapse rate*. Within each graph the area of symbols is proportional to the proportion of trials utilized at given stimulus intensity. Each graph also displays the generating $F(x; \alpha, \beta)$ for comparison

Fitting

All simulated data sets were fit using the following procedure:

1. Perform brute-force grid search. Likelihood values were computed across a two-dimensional grid across *location* and *slope* values. The grid contained all combinations of 601 *location* values that were equally spaced between $\alpha = F_{0.01}^{-1}$ and $\alpha = F_{0.99}^{-1}$, and 301 *slope* values that were equally-spaced on a log scale between -2 ($\beta = 0.01$) and 2 ($\beta = 100$). The *guess rate* was fixed at the value appropriate for the simulated procedure (i.e., $\gamma = 0.25$ or 0.5 for the 4AFC and 2AFC task, respectively). The *lapse rate* was fixed at $\lambda = 0.02$ (the generating value).
2. Perform Nelder-Mead (1965) iterative search. The PF with the highest likelihood in the brute-force grid search of step 1 served as the seed for the iterative search. The *guess rate* was fixed at the value appropriate for the simulated procedure (i.e., $\gamma = 0.25$ or 0.5 for the 4AFC and 2AFC task, respectively). The *lapse rate* was fixed at $\lambda = 0.02$ (the generating value). The search algorithm used was the Nelder-Mead (1965) Simplex search as implemented in the Palamedes toolbox (Prins & Kingdom, 2018, version 1.9.1).
3. Perform brute-force search through the likelihoods for all candidate Scenario = -1 fits. Likelihoods were calculated according to equation 4. Note that if k different stimulus intensities were used there are at most k possible candidates in this scenario (e.g., since the observed proportion correct for x_3 in Fig. 4 was below γ , there is no candidate model in this scenario for $s = 3$).
4. Perform brute-force search through the likelihoods for all candidate Scenario = -2 fits. Likelihoods were calculated according to equation 6. Note that if k different stimulus intensities were used there are $k - 1$ possible candidates in this scenario.
5. Calculate the likelihood for the single Scenario = -3 fit candidate according to equation 8.

Figure 7 displays the proportions of fits that fell into each of the scenarios. With the exception of very low N conditions, the rate with which a true maximum does not exist in the likelihood function is rather low. The pattern of results follows some predictable trends. An increase in the number of trials increases the probability that a true maximum exists in the likelihood function. When the $\Psi_{\alpha, \beta}$ method was used, all likelihood functions contained a true maximum with as few as 128 trials. The best-performing MOCS strategy on the other hand still has a low probability of lacking a true maximum in the likelihood function at $N = 128$ trials. I did not systematically search for the most optimal MOCS placement strategy, and it is likely that an MOCS placement strategy can be found that outperforms those I used here, if only slightly. On the other hand, it should be noted that in these simulations the generating function was known and that stimulus placements were centered on this known generating function. In real experiments the generating function will of course be unknown (otherwise there would be no point in conducting the experiment) and placement will generally not be optimally centered on the generating function.

Predictably, a placement strategy that covers a narrow range of stimulus intensities is more likely to result in Scenario = -3 (in which a constant function fits data better than any sigmoidal function) compared to a placement strategy that covers a wider range of stimulus intensities. On the other hand, a narrow placement strategy reduces the probability of Scenarios -1 and -2 (in which a step function fits data better than any sigmoidal function). Using a finer distribution of stimulus intensities (by increasing the number of stimulus intensities used from four to eight without increasing the total number of trials or changing the width of the range of stimulus intensities used) reduces the overall probability of the existence of a true maximum slightly, but does lead to an increase in the probability of Scenario -1 with a concomitant reduction in the probability of Scenario -2. It is also found, not surprisingly, that a true maximum is more likely to exist when a 4AFC, rather than a 2AFC design, is utilized.



Fig. 7. Proportion of fits resulting in each of the scenarios for the simulations in which a fixed *lapse rate* was used. Each ring corresponds to a different value of N (32, 64, 128, or 256 as labeled)

In all conditions tested, the adaptive $\text{Psi}_{\alpha\beta}$ method shows the highest probability of the existence of a true maximum in the likelihood function. When the stimulus placements guided by $\text{Psi}_{\alpha\beta}$ were used again but now in a MOCS design with a resampling of responses (“ $\text{Psi}_{\alpha\beta}$ resampled”), the rate of occurrence of Scenario = 1 (true global maximum exists) was comparable to the MOCS condition that displayed the lowest rate (“MOCS medium”). This indicates that it is the adaptive nature with respect to previous trials, rather than $\text{Psi}_{\alpha\beta}$ merely optimizing placement of stimuli with respect to the generating function, that is responsible for $\text{Psi}_{\alpha\beta}$'s superior performance.

While not a primary focus of this paper, for completeness, histograms of obtained parameter estimates are shown in Fig. 8. Median estimates are indicated by the vertical lines in the histograms. Note that allowing parameters to take on the “value” of $-\infty$ or $+\infty$ (as described above), even when no true maximum exists, allows for the determination of a median value across all simulations. *Location* parameter estimates display surprisingly little bias even at the lowest number of trials used, at least when judged by the median estimate. There is a moderate systematic bias apparent in the *slope* estimate when N is very low and a 2AFC procedure is utilized. It should be stressed, however, that in all these simulations the fixed value for the *lapse rate* corresponded to the true, generating value. In actual research the generating value for the *lapse rate* will be unknown. The effect of a mismatch between assumed and generating *lapse rate* on bias is not a topic of this current paper and has been noted and investigated elsewhere (e.g., Manny & Klein, 1985; Swanson & Birch, 1992; Prins, 2012).

Simulations using a free lapse rate

In the simulations described here the *lapse rate* was allowed to vary during the fitting process, but was constrained to lie in the interval $[0, 0.1]$. All simulations used the 2AFC design. Some of the previously used placement strategies were used here again and some additional placement strategies were added that were specifically geared towards providing information

regarding the value of the *lapse rate*. Placement strategies are displayed in Fig. 9. The previous MOCS placements using eight stimulus intensities (Fig. 6) were utilized here again. The $\text{Psi}_{\alpha\beta}$ placement was also used again. While these same placements were identical, now they were fitted while the *lapse rate* was free to vary.

Some additional placement strategies that are specifically geared towards obtaining a reliable estimate of the *lapse rate* were also used. I have previously demonstrated that, contrary to reports by Wichmann and Hill (2001), freeing the *lapse rate* will generally not eliminate biases in *location* and *slope* estimates (Prins, 2012). This was later confirmed by Linares and López-Moliner (2016). In the same paper I proposed an alternative strategy (joint Asymptotic Performance Lapse Estimation; “jAPLE”) that does essentially eliminate bias in *location* and *slope* parameters. In jAPLE, a stimulus intensity is included that is at an intensity that is so high that it may be assumed that an error at that intensity can only be due to a lapse. I refer to such an intensity as Asymptotic Performance Intensity or API. Critically, the model to be fitted includes the assumption that an incorrect response at an API can only have occurred due to a lapse. In other words, the model that is fit is:

$$\begin{aligned} \psi(x; \alpha, \beta, \gamma, \lambda) &= 1 - \lambda \text{ when } x = a \\ \psi(x; \alpha, \beta, \gamma, \lambda) &= \gamma + (1 - \gamma - \lambda) \\ &\quad \times F(x, \alpha, \beta, \gamma, \lambda) \text{ otherwise} \end{aligned} \quad (11)$$

In equation 11, stimulus intensity a is an API. Note that while observations made at $x = a$ contribute to the estimate of the *lapse rate* only, the *lapse rate* estimate is not solely determined by observations made at $x = a$. That is, observations at other stimulus intensities also contribute to the estimate of the *lapse rate*. In order to implement the jAPLE strategy here using an MOCS procedure, the medium width MOCS procedure was modified to include an API. Specifically, the original range of stimulus intensities was now covered by only seven of the eight intensities, while the eighth intensity was placed at an intensity of $+\infty$. Since $F(x = \infty; \alpha, \beta, \gamma, \lambda)$ evaluates to 1,

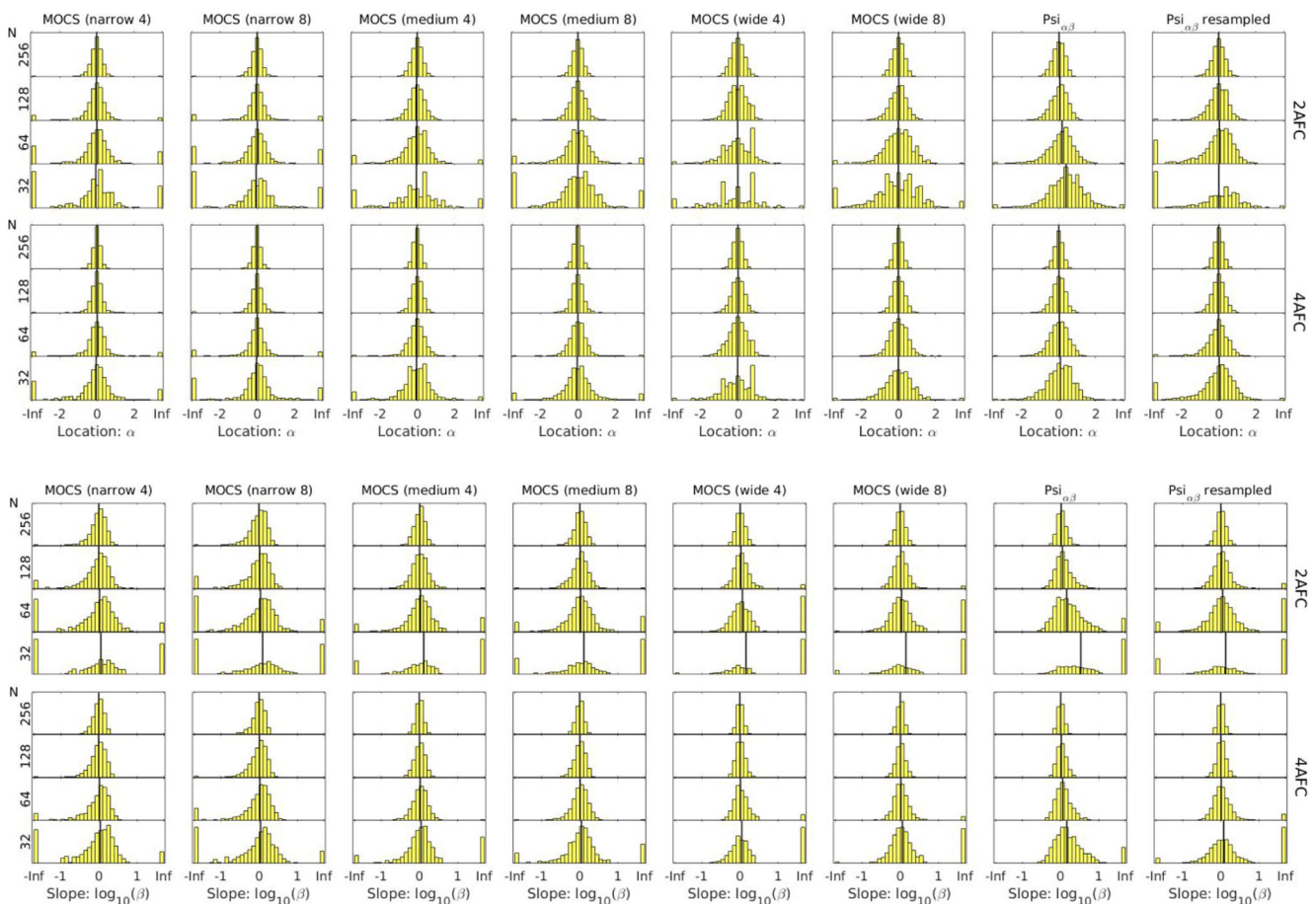


Fig. 8. Histograms of *location* and *slope* estimates for the fits in which a fixed *lapse rate* was used. Vertical lines indicate the location of the median in each distribution of parameter estimates. Median is based on all fits including those that did not result in Scenario 1

the simulated probability of a correct response at $x = +\infty$ equals $1 - \lambda$, and equation 11 is effectively being fitted to the data without additional modification to the Palamedes code. In Fig. 9 (and later figures), this condition is labeled “MOCS API jAPLE.” In order to separate the effect of using jAPLE over and above the mere inclusion of an API, another condition was added in which an (effective) API was included, but the model fitted was as before (i.e., equation 1). This was accomplished by including a stimulus intensity with a very high, but nevertheless finite, intensity (specifically $F_{0.99999}^{-1} = 11.51$). In the figures, this condition is labeled “MOCS API

nAPLE,” where nAPLE stands for non-Asymptotic Performance Lapse Estimation.

An additional placement strategy was the “psi-marginal” method (Prins, 2013). The psi-marginal method is similar to the original psi-method ($\psi_{\alpha\beta}$) except that the method maintains a posterior distribution across all free parameters (here: the *location*, the *slope* and the *lapse rate*) but selects stimulus intensities that will minimize the entropy in the posterior distribution in which any nuisance parameters (here: the *lapse rate*) has been marginalized. In the present context, this allows the method to utilize high

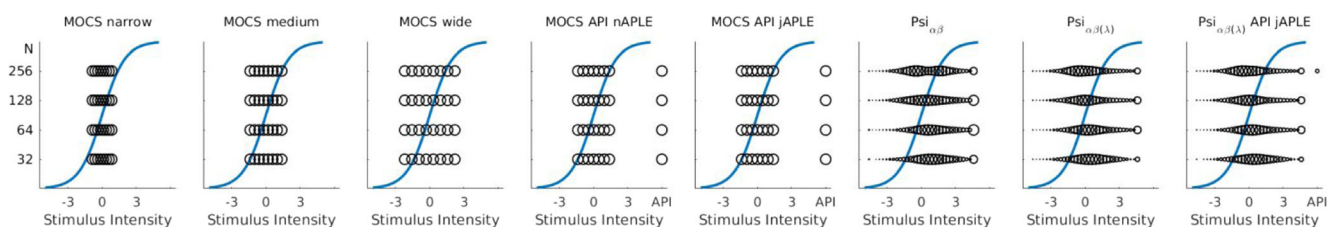


Fig. 9. Placements used in the simulations that were fit using a *free lapse rate*. Within each graph the area of symbols is proportional to the proportion of trials utilized at given stimulus intensity. Each graph also displays the generating $F(x; \alpha, \beta)$ for comparison

stimulus intensities (that would be informative for estimating the *lapse rate*) but will only do so if this is the optimal placement to reduce entropy regarding the values of the *location* and *slope* parameters. I label this condition as “ $\psi_{\alpha\beta(\lambda)}$ nAPLE”. Note that the *lapse rate* (λ) is now included in the subscript (because it is included in the posterior distribution) but is enclosed within parentheses to indicate it is marginalized before the expected entropy is calculated. The second modification to $\psi_{\alpha\beta}$ is identical to $\psi_{\alpha\beta(\lambda)}$ except that an API stimulus intensity is included and the likelihood function (and thus also posterior distribution) is calculated using equation 11. This can be effected simply in Palamedes' ψ method routines by including a stimulus intensity that is equal to $+\infty$ without necessitating other changes in the Palamedes code. This condition is labeled here as “ $\psi_{\alpha\beta(\lambda)}$ jAPLE.”

The rate at which each of the scenarios occurred for these simulations is displayed in Fig. 10. It is immediately obvious that allowing the *lapse rate* to vary greatly reduces the probability that the likelihood function will contain a true maximum. Even with $N = 256$ trials, likelihood functions are occasionally obtained that contain no true maximum for all of the placement/fitting regimes. It is clear that the MOCS placement strategies are much more susceptible to failed fits compared to the adaptive methods, though it should be noted also that the inclusion of an API brings the performance of the MOCS placements near that of the $\psi_{\alpha\beta}$ and $\psi_{\alpha\beta(\lambda)}$ methods. Interestingly, the condition “ $\psi_{\alpha\beta(\lambda)}$ API jAPLE” leads to more likelihood functions lacking a true maximum compared to the $\psi_{\alpha\beta(\lambda)}$ method. With regard to avoiding a lack of a true maximum in the likelihood function, $\psi_{\alpha\beta(\lambda)}$ performs best among the methods considered.

While not a primary focus of this study, for completeness Fig. 11 displays histograms of all parameter estimates that could be assigned a value as well as the median parameter (where it could be determined; remember that if Scenario = -3 and the *lapse rate* is free, the *location* and *lapse rate* parameter show complete redundancy and cannot reasonably be assigned a value even if we allow assignment of $+\infty$ or $-\infty$ values). In the conditions where the median can be determined, the *location* parameter is relatively bias-free. However, the *slope* parameter displays systematic and large bias in most conditions with the exception of conditions in which the highest value of N was used.

Discussion

In order to fit a psychometric function (PF) using a maximum-likelihood criterion, the maximum in the likelihood function must be located. This generally cannot be accomplished using analytical methods and must instead be performed using numerical approximation. This procedure is susceptible to failure. The fitting procedure may mistake a local maximum as the global maximum. When the likelihood function does not contain a (global) maximum, the procedure may nevertheless incorrectly report that a maximum in the likelihood function was found. Even in simple designs with an even distribution of trials across few stimulus intensities, it may be difficult to detect such false fits (e.g., Fig. 2). In designs with uneven distribution of trials or in which each trial uses unique stimulus intensities, as may occur in adaptive methods, detecting false fits becomes increasingly difficult. Moreover, when false fits occur in bootstrap simulations in order to determine reliability of parameter estimates, to determine Goodness-of-fit or to perform model comparisons results will be inaccurate, possibly resulting in incorrect conclusions being drawn from the experiment.

Here I have proposed and tested a method that may be used not only to avoid convergence on a local maximum in the likelihood function, but also to determine whether a true maximum in the likelihood function actually exists. In case a maximum does not exist in the likelihood function, the method identifies which of three possible scenarios describes the data best. In each of the three scenarios, a limiting function that may be approximated to any degree of precision by a sigmoidal PF (but which cannot match the limiting function exactly) is associated with a higher likelihood than any sigmoidal PF.

The simulations presented here investigated the probability with which the likelihood function lacks a true maximum as a function of total number of trials, the number of response alternatives in an mAFC design, the stimulus placement strategy, and whether the *lapse rate* was fixed or was allowed to vary. It was found that many experimental designs (including the number of trials used) that would be considered reasonable still resulted in likelihood functions that lack a maximum with a relatively high rate. Even adaptive placement strategies that can reasonably be considered near optimal resulted in the lack of a maximum in the likelihood function for a significant proportion of simulated experiments.

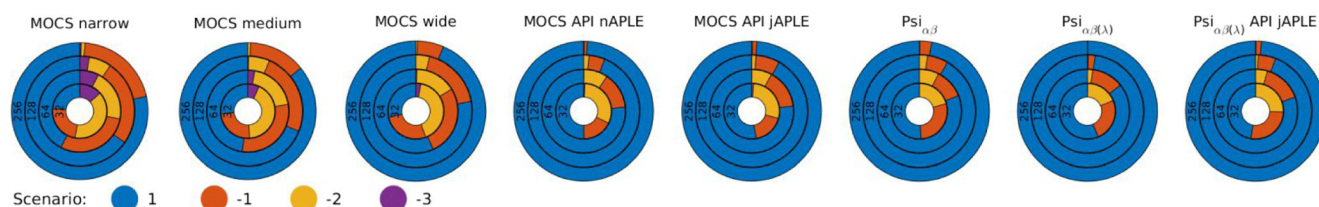


Fig. 10. Proportion of fits resulting in each of the scenarios for the simulations in which a free *lapse rate* was used

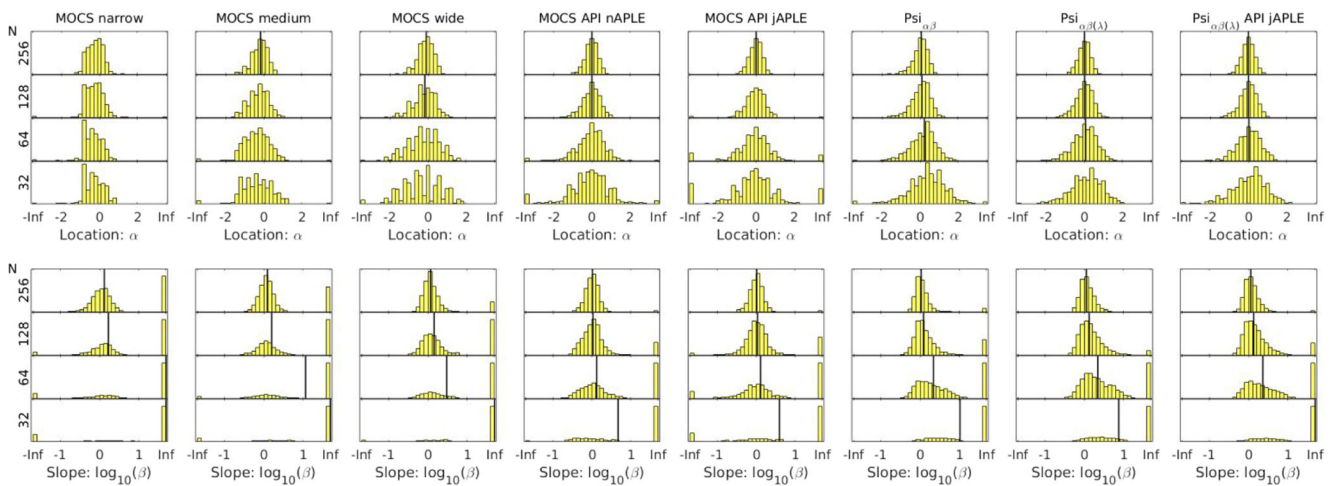


Fig. 11. Histograms of *location* and *slope* estimates for the fits in which a free *lapse rate* was used. Vertical lines indicate the location of the median in each distribution of parameter estimates. Median is based on all fits

Note that the simulations presented here all represent somewhat of a best-case scenario that one cannot expect to match in real experiments. All of the assumptions that the MOCS and adaptive placements methods made were omniscient with respect to the generating function. For example, all MOCS placement strategies used here were positioned relative to the generating function. In actual research the *location* and *slope* parameter values and the form (Logistic, Weibull, etc.) of the generating function will be unknown, and any MOCS placement strategy will not be aligned as well with it as was the case here. Placement by the adaptive $\psi_{\alpha\beta}$ method relied much less on omniscience regarding *location* and *slope* parameters but did use the true, generating value of the *lapse rate*. Finally, all maximum-likelihood fits were omniscient with respect to the generating function in the assumptions that they made. For example, all assumed the true generating form of the PF. Those that assumed a fixed *lapse rate* assumed the correct, generating *lapse rate*.

The core problem underlying failed fits is that of overspecification of the model. Whenever a fit fails because no maximum in the likelihood function exists, it means that the model is overspecified: it includes variables that the data do not contain sufficient information on to support their estimates. Overspecification of a model allows the model to accommodate sampling error that is not informative of the underlying process. Inclusion of the *lapse rate* is especially problematic when it comes to model overspecification since the *lapse rate* can be highly redundant with the *location* and *slope* parameters. The degree of redundancy is mainly dependent on the stimulus placement and the number of trials (e.g., Prins, 2012). Especially when stimulus placements are concentrated around the value of the *location* parameter (i.e., as in the “narrow” MOCS placements used here) is the *lapse rate* highly redundant with the *location* and *slope* parameters. In effect, what this means is that the value of the *lapse rate* parameter

including those that did not result in Scenario 1. Note that a median *location* is not shown in all conditions since no value (finite or infinite) can be assigned to *location* when Scenario = -3 and the *lapse rate* is free

can trade off with those of the *location* and *slope* parameter to result in similar predicted probabilities of correct response within the limited range of stimulus intensities used. Effectively, there are multiple combinations of parameter values that may be consistent with the data. This leads to the existence of multiple regions of high likelihood in the likelihood function (including perhaps a region consistent with one of the scenarios discussed here). Indeed, the rise in the probability of failed fits when the *lapse rate* is freed observed in the simulations is especially pronounced in the “narrow” MOCS placement, followed by the “medium,” then the “wide” MOCS placement.

It is important to stress that the ability offered here to identify a step or constant function that has a higher likelihood function than a strictly sigmoidal function is not intended to remedy a fit that would otherwise fail. Any fit that results in scenario -1, -2, or -3 should still be regarded as a failed fit and instead should be taken to indicate that the model that is being fitted is overspecified or perhaps that the experiment was poorly designed, specifically with regard to the placement of stimulus intensities.

In order to avoid failed fits, some specific guidelines for the planning of an experiment and the selection of model to be fitted to the resulting data may be derived from the results presented here. Most importantly, perhaps, it is key to realize that the complexity of the model should match the availability of information in the data in order to avoid overspecification of the model. One should not free a parameter simply because one can. As noted, this is especially a concern with the *lapse rate* because of its redundancy with the PF's other parameters. Moreover, as has been shown before (Linares & López-Moliner, 2016; Prins, 2012), unless the design supports the estimation of a *lapse rate* (i.e., contains sufficient information regarding the value of the *lapse rate*), allowing the *lapse rate* to vary does little to improve bias in the parameters of interest.

It should be stressed, however, that failing to allow at all for the occasional lapse in attention or finger error by utilizing a model that assumes a *lapse rate* of zero may lead to severe bias in *location* and *slope* parameters if a lapse does in fact occur. This has been noted as early as 1981 by Hall, and many times since (e.g., Linares & Lopez-Moliner, 2016; May & Solomon, 2013; Manny & Klein, 1985; Prins, 2012; Swanson & Birch, 1992; Treutwein & Strasburger, 1999; Wichmann & Hill, 2001). Thus, if one chooses to fix the lapse rate at a specific value, this value should be greater than zero.

Regardless of the complexity of the model, not surprisingly, the more trials one uses, the less likely it is that the likelihood function lacks a maximum. Use of a larger number of trials may not always be an option, especially perhaps in clinical settings, with some special populations, or when testing requires a large amount of resources, in which case one should use as simple a model as possible in order to avoid overspecification.

Stimulus placement affects the probability of occurrence of the three scenarios in predictable ways, thus it is important to plan stimulus placement carefully. The narrower the range in which the stimuli are placed, the closer to each other the true probabilities of a correct response will be and the likelier it is that there is no overall rising trend in the proportion correct with increasing stimulus intensity. In such a case, a constant function will fit better than a strict sigmoidal function (Scenario -3). Likewise, spreading the stimuli across a wider range will increase the likelihood that a step function will provide a better fit than a strict sigmoidal function (Scenarios -1 and -2). The use of an adaptive method is advised as it resulted in the lowest rate of occurrence of a lack of a maximum in the likelihood function. When one uses an adaptive method one should not free the *lapse rate* parameter in the fit unless the adaptive method targeted the *lapse rate*. Finally, when the experimental design allows it, one should use a high number of alternatives in an mAFC task.

Unless the *slope* value is of specific theoretical interest, one has the option to fix it at a reasonable value. This will ensure that the likelihood function contains a true global maximum. Alternatively, if a participant is tested in multiple conditions and it can be assumed that the *slope* value is constant across conditions, one can estimate a single *slope* value across the conditions (while allowing the *location* parameter to vary between conditions). This will have a similar effect to increasing the number of trials.

Finally, one may consider a Bayesian approach to fitting a PF (e.g., Kingdom & Prins, 2016; Kuss et al., 2005; Schütt et al., 2016). The Bayesian approach allows one to view experimental results as merely modifying one's prior beliefs regarding parameter values. Thus, final parameter estimates are based not only on experimental results (as is the case with maximum-likelihood estimation), but also on one's prior beliefs regarding likely parameter values. However, while the proposal that a prior across *slope* values should not include

0 (Scenario -3) or infinity (Scenarios -1 and -2) should find little resistance in and of itself, it may be difficult to conceive of a full prior distribution across *slope* values that researchers and consumers of research will universally agree with (but see Schütt et al., 2016).

As of Palamedes version 1.9.1, the routines in the Palamedes toolbox (Prins & Kingdom, 2018) that use a maximum-likelihood criterion to fit individual PFs incorporate the method described above by default and will report whether a true maximum in the likelihood function exists and if not, which scenario best describes the experimental results.⁴

Acknowledgements Some of the results presented here were also presented as a poster at the 17th Annual Meeting of the Vision Sciences Society. Prins, N. (2017) *Journal of Vision*, 17(10): 788-788.

Compliance with ethical standards

Open practices statement All reported simulations and psychometric fits were performed using the Palamedes Toolbox version 1.9.1 available at www.palamedestoolbox.org. No experiments were preregistered.

Appendix

By a well-known theorem any continuous function defined across a bounded (i.e., finite) and closed (i.e., bounds are included) domain S defined in \mathbb{R}^n will attain a finite maximum value within S (e.g., Williamson & Trotter, 1996; p. 221). While in the present context the likelihood function is continuous within its domain, its domain is not bounded and closed and this is the reason that the likelihood function may lack a maximum. The domain for the *location* parameter is either all real numbers (e.g., Logistic) or all non-negative real numbers (e.g., Weibull). The domain for the *slope* parameter is all non-negative real numbers. The domain for both the *guess* and *lapse rate* parameters cannot exceed $[0, 1]$. Continuous functions for which the theorem's stipulations on the domain are not met may lack a finite maximum (or minimum) in two general circumstances. One, the function's range may extend to $-\infty$ or $+\infty$. This is the reason that, for example, the function $f(x) = x^{-1}$ does not have a (finite) maximum in the domain $(0, 1)$ (note that this example exemplifies why the theorem not only stipulates that the domain must be bound but also must be closed). This cannot be the reason for a lack of a maximum in the likelihood function, however, since any likelihood function will have a limited range (likelihood must by definition be within $[0, 1]$). The second circumstance under which a continuous function may lack a maximum is when the function asymptotes toward a value (that is higher than any other function value within its domain) when one or more of the variables in its domain approaches $-\infty$ or $+\infty$. This further necessitates the theorem's

⁴ The Palamedes Toolbox can be downloaded at www.palamedestoolbox.org

stipulation that the domain is bound. This circumstance prevents, for example, the function $f(x) = -x^{-1}$ having a maximum in the domain $(0, +\infty)$. Taken together, the above implies that, in the context of maximum-likelihood fitting of PFs a lack of a maximum can only occur when an asymptotic likelihood value (that is higher than any other likelihood value in the domain) is approached as the *location* parameter approaches $-\infty$ or $+\infty$ or as the *slope* parameter approaches $+\infty$. The PF will then approach either a step function (as the *slope* parameter approaches $+\infty$, for any value of the *location* parameter) or it will approach, within the range of stimulus intensities used, a constant function at any height within $[\gamma, 1 - \lambda]$ (as the *location* parameter approaches $-\infty$ or $+\infty$ while the *slope* parameter co-varies to approach any particular height of the function $[\gamma, 1 - \lambda]$ within the range of stimulus intensities used). Thus, the lack of existence of a maximum in the likelihood function implies that either an increasing step function or a constant function (both of which can be approached by a sigmoidal function within a limited range of stimulus intensities) has a likelihood that is higher than that of the sigmoidal function but can be approached by it.

References

- Albert, A., & Anderson, J.A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, 71(1), 1–10.
- Efron, B. & Tibshirani, R.J. (1993). An introduction to the bootstrap. Chapman & Hall/CRC, Boca Raton, FL.
- Fründ, I., Haenel, N.V., & Wichmann, F.A. (2011). Inference for psychometric functions in the presence of nonstationary behavior. *Journal of Vision*, 11(6), 16, 1–19.
- Hall, J.L. (1981). Hybrid adaptive procedure for estimation of psychometric functions. *Journal of the Acoustical Society of America*, 69(6), 1763–1769.
- Kingdom, F.A.A. & Prins, N. (2016). *Psychophysics: A Practical Introduction*, 2nd Ed. Academic Press, San Diego, CA.
- Kolda, T.G., Lewis, R.M., & Torczon, V. (2003). Optimization by direct search: new perspectives on some classical and modern methods. *SIAM Rev.* 45, 385–482.
- Kontsevich, L.L. & Tyler, C.W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, 39, 2729–2737.
- Kuss, M., Jäkel, F., & Wichmann, F.A. (2005). Bayesian inference for psychometric functions. *Journal of Vision*, 5, 478–492.
- Lagarias, J.C., Reeds, J.A., Wright, M.H., & Wright, P.E. (1998). Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM J Optim.* 9(1), 112–147.
- Linares, D. & López-Moliner, J. (2016). quickpsy: An R Package to Fit Psychometric Functions for Multiple Groups. *The R Journal*, 8:1, 122–131.
- Manny, R.E. & Klein, S.A. (1985). A three alternative tracking paradigm to measure Vernier acuity of older infants. *Vision Research*, 25(9), 1245–1252.
- May K. A., Solomon J. A. (2013). Four theorems on the psychometric function. *PLoS One*, 8 (10), e74815.
- Nelder, J.A. & Mead, R. (1965). A simplex method for function minimization. *Comput. J.* 7, 308–313.
- Pratt, J.W. (1981). Concavity of the Log Likelihood. *Journal of the American Statistical Association*, 76(373), 103–106.
- Prins, N. (2012). The psychometric function: The lapse rate revisited. *Journal of Vision*, 12(6): 25, 1–16.
- Prins, N. (2013). The psi-marginal adaptive method: How to give nuisance parameters the attention they deserve (no more, no less). *Journal of Vision*, 13(7): 3, 1–17.
- Prins, N. & Kingdom, F.A.A. (2018). Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes Toolbox. *Frontiers in Psychology*, 9: 1250.
- Schütt, H.H., Harmeling, S., Macke, J.H., & Wichmann, F.A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Research*, 122, 105–123.
- Swanson, W.H. & Birch, E.E. (1992). Extracting thresholds from noisy psychophysical data. *Perception & Psychophysics*, 51(5), 409–422.
- Swets, J.A. (1961). Is there a sensory threshold? *Science*, 134, 168–177.
- Treutwein, B. & Strasburger, H. (1999). Fitting the psychometric function. *Perception & Psychophysics*, 61(1), 87–106.
- van Driel, J., Knapen, T., van Es, D.M., & Cohen, M.X. (2014). Interregional alpha-band synchrony supports temporal cross-modal integration. *Neuroimage*, 101, 404–415.
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. fitting, sampling, and goodness of fit. *Perception and Psychophysics*, 63(8), 1293–1313.
- Williamson, R.E. & Trotter, H.F. (1996). *Multivariable Mathematics*, 3rd Ed. Prentice Hall: Upper Saddle River, NJ

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.